

INSTRUMENTOS PARA LA GESTIÓN DE PROYECTOS DE EXPLOTACIÓN DE INFORMACIÓN

García-Martínez, R., Merlino, H., Diez, E., Rodríguez, D., Pytel, P.,
Baldizzoni, E., Arboleya, H., Martins, S., Cartanilica, A.

Grupo Investigación en Sistemas de Información

Departamento de Desarrollo Productivo y Tecnológico. Universidad Nacional de Lanús
29 de Septiembre 3901 (1826) Remedios de Escalada, Lanús. Argentina. Tel +54 11 6322-9200 Ext. 194
rgarcia@unla.edu.ar

RESUMEN

Los Proyectos de Explotación de Información difieren sustancialmente de los de Software convencional. Las herramientas involucradas en los procesos de Ingeniería en Software no son aplicables a este tipo de proyectos. Surge la necesidad de desarrollar un cuerpo de conocimientos para una Ingeniería de Explotación de Información, cuyo eje se centre en el desarrollo de instrumentos que permitan un adecuado control y gestión de los proyectos, contribuyendo a mejorar la productividad de las PYMES del área.

Palabras clave: Proyectos de explotación de información, Instrumentos de gestión, test de viabilidad, modelo de ciclo de vida, mapa de actividades, formalismos para documentación de requisitos, método de estimación de recursos.

CONTEXTO

Durante la etapa de investigación documental de los proyectos UNLa 33A81 y UNLa 33A105 desarrollados por el grupo de trabajo, se observó que en la bibliografía consultada se repetía con insistencia el uso indiscriminado de los términos “minería de datos” (o data mining) y “explotación de información” (o information mining) para referirse al mismo cuerpo de conocimientos. En el marco de este proyecto, se considera que la minería de datos se refiere a la algoritmia para encontrar patrones de conocimiento en masas de datos; mientras que la ingeniería de explotación de información entiende en los procesos y las metodologías utilizadas para ordenar, controlar y gestionar la tarea de encontrar patrones de conocimiento referidos (García-Martínez et al. 2011a).

Los proyectos de explotación de información poseen características muy distintas a los proyectos de desarrollo de software tradicional (Pollo-Cattaneo et al., 2010a), sobre todo en la parte operativa del proyecto. La diferencia se presenta en los procesos de desarrollo y mantenimiento en los cuales el ciclo

de fases de un proyecto de software tradicional: inicio, requisitos, análisis y diseño, construcción, integración y pruebas no resultan naturales en un proyecto de explotación de información (Vanrell et al., 2010a; 2010b; 2012). Por otra parte, al evaluar las principales metodologías existentes para los proyectos de explotación de información (Chapman et al., 1999; SAS, 2008; Pyle, 2003), se observa la falta de herramientas que permitan soportar de forma completa la fase de administración de proyectos (Vanrell 2012).

Durante el desarrollo el proyecto 33A105 “Ingeniería de Proyectos de Explotación de Información para PYMES” se pudo producir un primer ordenamiento del cuerpo de conocimiento existente sobre lo que se ha definido como línea de investigación en Ingeniería de Explotación de Información. Una de las áreas de vacancia identificadas es la de instrumentos que permitan un adecuado control y gestión de proyectos de explotación de información (García-Martínez et al., 2011c).

Una de las lecciones aprendidas sobre desarrollo de software en Informática derivada de los estadios tempranos de la disciplina, es que, la ausencia de una ingeniería de software conllevaba a un desarrollo artesanal de los artefactos software (Ochoa et al., 2008). El desarrollo artesanal implicaba la imposibilidad de poder establecer dentro de valores racionales, parámetros tales como: [a] cantidad y calificación de los recursos humanos a emplear en el proyecto, [b] tiempos de desarrollo, [c] modelos de proceso que guiaran el desarrollo y permitieran establecer hitos de entrega, [d] formalismos de documentación que dieran cuenta de lo hecho en el proyecto de desarrollo del artefacto software y de las decisiones de diseño asumidas, constituyendo el punto de partida para futuras ampliación de funcionalidades, [e] modelos de costo de proyecto (Boehm, 1981), entre otros. De hecho la estimación de estos parámetros se hacía en base a la experiencia de individuos sin ninguna base ingenieril y lo para un grupo de desarrollo podía hacerse en meses para otro podía hacerse en años.

La explotación de información esta en sus primeros estadios y, al igual que lo que ocurría con el desarrollo de artefactos software, adolece de una ingeniería que provea instrumentos para un adecuado control y gestión de proyectos de este tipo. En particular, en el caso de las PYMES del sector, reviste importancia por cuanto contar con herramientas para el control y la gestión de proyectos de explotación de información les permite explorar una nueva línea de servicios informáticos. Disponer de modelos de proceso para este tipo de proyectos, habilita a la PYME disponer de una guía que ordene el trabajo interno y permita establecer y ofrecer al requirente hitos para los entregables del proyecto. Disponer de modelos que permitan cuantificar: la viabilidad, los recursos y los tiempos a emplear en el proyecto de explotación de información, hace predecibles (y acotables) los costos al momento que una PyME ofrezca este tipo de servicios en el marco de aplicaciones gubernamentales o comerciales. Por otra parte disponer de formalismos de documentación facilita a la PYME repetir procesos con el consecuente ahorro de recursos.

INTRODUCCIÓN

Hace ya un lustro, un estudio de la Universidad de California en Berkeley [Lyman y Varian, 2003] señaló que la información disponible en Internet crecía a razón de 92 petabytes [1015 bytes] por año. En [Maimon y Rokach, 2005] se ha señalado que esta información esta disponible para procesos de descubrimiento de conocimiento con independencia que se encuentre en fuentes estructuradas [Rudin y Cressy, 2003; Moss, 2003] ó desestructuradas [Vuori, 2006].

La inteligencia de negocio propone un abordaje interdisciplinario [dentro del que se encuentra la Informática], que tomando todos los recursos de información disponibles y el uso de herramientas analíticas y de síntesis con capacidad de transformar la información en conocimiento, se centra en generar a partir de éstos, conocimiento que contribuya con la toma de decisiones de gestión y generación de planes estratégicos en las organizaciones [Thomsen, 2003; Negash y Gray, 2008].

La Explotación de Información es la sub-disciplina de los Sistemas de Información que aporta a la Inteligencia de Negocio [Langseth y Vivatrat, 2003] las herramientas para la transformación de información en conocimiento [Mobasher et al., 1999; Srivastava et al., 2000; Abraham, 2003; Coley, 2003]. Ha sido definido como la búsqueda de

patrones interesantes y de regularidades importantes en grandes masas de información [Fayad et al., 1996; Grossman et al., 1998].

Un Proceso de Explotación de Información se define, como un grupo de tareas relacionadas lógicamente [Curtis et al., 1992] que, a partir de un conjunto de información con un cierto grado de valor para la organización, se ejecuta para lograr otro, con un grado de valor mayor que el inicial [Ferreira et al., 2005; Kanungo, 2005]. Adicionalmente, existe una variedad de técnicas de minería de datos, en su mayoría provenientes del campo del Aprendizaje Automático [García-Martínez, 1997; García-Martínez et al., 2003], susceptibles de ser utilizadas en cada uno de estos procesos.

El proyecto que se presenta es continuación del Proyecto UNLa 33A105 “Ingeniería de Proyectos de Explotación de Información para PyMEs” [05/2011 – 04/2013], que fue a su vez continuación del Proyecto UNLa 33A081 “Sistemas de Información para Inteligencia de Negocio” [05/2009 – 04/2011]. Los resultados de esto proyectos son:

- En el área de Fundamentos se ha argumentado la necesidad de desarrollar una Ingeniería de Proyectos de Explotación de Información para PYMES [García-Martínez et al., 2011c]; identificando las áreas prioritarias sobre las cuales trabajar en PyMEs [García-Martínez et al., 2010a].
- En el área de Procesos para Proyectos de Explotación de Información [EI], se ha formulado una propuesta de procesos de EI [Britos y García-Martínez, 2009; Rancan et al., 2010; García-Martínez et al., 2011b; Pollo Cattaneo et al., 2012a], se ha fundamentado la necesidad de definir técnicas para el aseguramiento de la calidad [Diez, et al., 2012], se ha realizado una caracterización empírica de dominios para uso en proyectos de EI [Lopez-Nocera et al., 2011]; se ha trabajado en desarrollar técnicas y una metodología específica para elicitación de requerimientos [Pollo Cattaneo et al., 2010b; Mansilla et al., 2012] identificando posibles formalismos de documentación [Vegega et al., 2012], se han sentado las bases para trabajar en viabilidad y estimación de proyectos de EI [Rodríguez et al., 2010; Pytel et al., 2011a; 2011b; 2011c; 2012], y se ha desarrollado un modelo de proceso de operación para proyectos de EI [Vanrell et al., 2010a; 2010b].
- En el área de Trabajo de Campo se validaron los resultados parciales obtenidos en problemas concretos en los dominios de: tuberías de gas [D’Atri et al., 2009], prevención de daños y averías en la industria automotriz [Flores et al.,

2009], caracterización de problemas de aprendizaje [Jiménez Rey et al., 2009], identificación de patrones de comportamiento de comunidades educativas mediadas por entornos virtuales [Cigliuti et al., 2012], identificación de errores de apropiación de conceptos en el dominio de análisis de sistemas de información [Saavedra-Martínez et al., 2012b; 2012b], prevención de estrés de suelos [Sanson et al., 2009], identificación de datos faltantes con ruido o inconsistentes en auditoria de sistemas [Kuna et al., 2010a; 2010b; 2011] realizando estudios comparativos de distintos procedimientos [Kuna et al., 2012], identificación de causales de abandono de estudios universitarios [Kuna et al., 2009; 2010c], y estimación en gestión de proyectos de software [Bogado, et al., 2011]. Previamente se trabajo en: identificación de caras humanas [Britos et al., 2005], detección de cambios de consumo de usuarios [Grosser et al., 2005; Britos et al., 2008], localización de patrones en eventos meteorológicos [Cogliati et al., 2006a; 2006b], predicción de la salud de una comunidad [Felgaer et al., 2006], detección de daños al corazón [Ferrero et al., 2006], registro de uso de sitios web [Britos et al., 2007], selección de protocolos pedagógicos [Britos et al., 2008a], comprobación de malentendidos en programación [Britos et al., 2008c], y detección de patrones criminales [Valenga et al., 2007a; 2007b; 2008].

OBJETIVOS E HIPOTESIS DE INVESTIGACION

En este proyecto se busca continuar con el desarrollo y sistematización del cuerpo de conocimiento asociado a la Ingeniería de Explotación de Información iniciado en el proyecto 33A105. En esta etapa, la investigación se focalizará en el desarrollo de herramientas para el Control y la Gestión de Proyectos de Explotación de Información en PyMEs.

La pregunta problema que anima la investigación es: ¿Se puede cubrir la vacancia de herramientas de control y gestión para proyectos de explotación de información?

Entre los supuestos (o hipótesis) que guían el proyecto se encuentran:

Hipótesis I: Los proyectos de explotación de información poseen características muy distintas a las de los proyectos de desarrollo de software tradicionales. Las clásicas etapas de análisis, diseño, desarrollo, integración y testeo no encajan con las etapas naturales de los procesos de desarrollo de este tipo de proyectos. En consecuencia, los

métodos de determinación de viabilidad de proyecto, los modelos ciclos de vida para software (p.ej.: cascada, prototipado, ó espiral); los modelos de procesos software (p.ej.: IEEE 1074 o MOPROSOFT); y la natural derivación de estos últimos: los mapas de actividades para proyectos software no son aplicables a este tipo de proyectos.

Hipótesis II: Existen metodologías de explotación de información que destacan la importancia del planeamiento de una elicitación de requerimientos a lo largo de todo el proyecto de una manera ordenada, documentada, consistente y trazable. Sin embargo, el abordaje clásico de la ingeniería de requerimientos no se ajusta a los proyectos de explotación de información porque no atiende los aspectos particulares de especificación de requerimientos para este tipo de proyectos. De hecho, las técnicas clásicas no son aplicables al proceso de identificar el problema de explotación de información ni la documentación asociada.

Hipótesis III: El proceso de aseguramiento de la calidad es el conjunto de acciones planificadas y sistemáticas implantadas dentro del sistema de calidad, y demostrables si es necesario, para proporcionar la confianza adecuada que una entidad cumplirá los requisitos de calidad de procesos y de productos en el marco de un proyecto de desarrollo. La Ingeniería de Explotación de Información dispone de metodologías que guían el desenvolvimiento de proyectos que se consideran probadas y tienen un buen nivel de madurez en cuanto a los procesos de desarrollo. Sin embargo, estas metodologías no definen procesos específicos para el control y la gestión de proyectos en el área, en particular, para el aseguramiento de la calidad.

Hipótesis IV: La determinación de las tareas a desarrollar en un proyecto, de los recursos necesarios y del tiempo que transcurrirá desde el comienzo hasta el final de su realización; es una actividad de control y gestión que permite definir tempranamente la duración del proyecto y formular el correspondiente cálculo de costos. Si bien el metodo DMCoMo permite estimar proyectos de explotación de información, sus autores han señalado su aplicabilidad solo para proyectos de grandes organizaciones. Esto deja a las PyMEs sin un procedimiento de estimación de proyectos de explotación de información,

Objetivo General: El objetivo de este proyecto es desarrollar el cuerpo de conocimiento necesario para las actividades tempranas de control y gestión en un proyecto de explotación de información con focalización en su transferencia al sector PyMEs de informática. Se busca proveer las siguientes

herramientas para proyectos de explotación de información: test de viabilidad, modelo de ciclo de vida, mapa de actividades, procedimiento de derivación del problema de negocio en problema de explotación de información, formalismos para documentación de requisitos, técnicas y métodos de aseguramiento de la calidad del proceso y del producto y método de estimación de recursos.

Objetivos específicos vinculados a Hipótesis I:

- 1.- Desarrollar de un Test de Viabilidad de Proyectos de Explotación de Información
- 2.- Desarrollar un Modelo de Ciclo de Vida para Proyectos de Explotación de Información
- 3.- Desarrollar un Mapa de Actividades para para Proyectos de Explotación de Información

Objetivos específicos vinculados a la Hipótesis II:

- 4.- Desarrollar un procedimiento basado en formalismos de Ingeniería del Conocimiento que permita derivar el problema de explotación de información a partir de representaciones del problema de negocio.
- 5.- Desarrollar formalismos basados en Ingeniería del Conocimiento que permitan la documentación de requisitos en proyectos de explotación de información

Objetivo específico vinculado a la Hipótesis III:

- 6.- Desarrollar técnicas y métodos para aseguramiento de la calidad del proceso y del producto en Proyectos de Explotación de Información.

Objetivo específico vinculado a la Hipótesis IV:

- 7.- Desarrollo de un Método de Estimación de Recursos para Proyectos de Explotación de Información en PyMES.

METODOLOGÍA DE TRABAJO

Se prevé: realizar investigación documental identificando casos de estudio, desarrollar mediante la metodología de prototipado evolutivo las versiones iniciales de los instrumentos combinado con mejora incremental, probar la versión de producción de los prototipos en casos seleccionados.

RESULTADOS OBTENIDOS/ESPERADOS

Como resultado de este proyecto, se esperan contar al fin de su desarrollo con los siguientes instrumentos para la gestión de proyectos de explotación de información: test de viabilidad, modelo de ciclo de vida, mapa de actividades, procedimiento de derivación del problema de negocio en problema de explotación de información, formalismos para documentación de requisitos,

técnicas y métodos de aseguramiento de la calidad del proceso y del producto, y método de estimación de recursos.

FORMACIÓN DE RECURSOS HUMANOS

El grupo de trabajo se encuentra formado por tres investigadores formados, dos investigadores en formación y cuatro becarios alumnos de la carrera Licenciatura en Sistemas de la UNLa. En su marco se desarrolla dos Tesis de Doctorado en Ciencias Informáticas y cuatro Trabajos Finales de Licenciatura en Sistemas.

REFERENCIAS

- Abraham, A. 2003. Business Intelligence from Web Usage Mining. Journal of Information & Knowledge Management, 24: 375-390.
- Böehm, B. 1981. Software Engineering Economics. Prentice Hall. ISBN 0-13-822122-7.
- Böehm, B. 1988. A Spiral Model of Software Development and Enhancement. IEEE Computer. Mayo 1988. Pág. 61-72.
- Bogado, V., Dapozo, G., García-Martínez, R. 2011. Estimación en Gestión de Proyectos de Software Basada en Explotación de Información. Proceedings XIII Workshop de Investigadores en Ciencias de la Computación. Artículo 3894.
- Britos, P., Abasolo, M., García-Martínez, R. y Perales, F. 2005. Identification of MPEG-4 Patterns in Human Faces Using Data Mining Techniques. Proceedings 13th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2005. Páginas 9-10.
- Britos, P., Cataldi, Z., Sierra, E., García-Martínez, R. 2008a. Pedagogical Protocols Selection Automatic Assistance. Notes in Artificial Intelligence 5027: 331-336.
- Britos, P., García-Martínez, R. 2009. Propuesta de Procesos de Explotación de Información. Proceedings XV Congreso Argentino de Ciencias de la Computación. Workshop de Base de Datos y Minería de Datos. Págs. 1041-1050. ISBN 978-897-24068-4-1.
- Britos, P., Grosser, H., Rodríguez, D., García-Martínez, R. 2008b. Detecting Unusual Changes of Users Consumption. In Artificial Intelligence in Theory and Practice II, ed. M. Bramer, Boston: Springer, 276: 297-306.
- Britos, P., Jiménez Rey, E., García-Martínez, R. 2008c. Work in Progress: Programming Misunderstandings Discovering Process Based On Intelligent Data Mining Tools. Proceedings 38th ASEE/IEEE Frontiers in Education Conference.
- Britos, P., Martinelli, D., Merlino, H., García-Martínez, R. 2007. Web Usage Mining Using Self Organized Maps. International Journal of Computer Science and Network Security, 76:45-50. ISSN : 1738-7906.
- Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 1999. CRISP-DM 1.0 Step by step Bgguide. Edited by SPSS. <http://www.crispdm.org/CRISPWP-0800.pdf>. Último acceso Junio 2008.
- Cigliuti, P., Pollo-Cattaneo, F., García-Martínez, R. 2012. Procesos de Identificación de Comportamiento de Comunidades Educativas Centradas en EVEAs. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 954-956. ISBN 978-950-766-082-5.
- Cogliati, M., Britos, P., García-Martínez, R. 2006a. Patterns in Temporal Series of Meteorological Variables Using SOM & TDIDT. In IFIP International Federation for Information Processing. Volume 217, Artificial Intelligence in Theory and Practice, ed. M. Bramer, Boston: Springer, Pág. 305-314.
- Cogliati, M., Britos, P., García-Martínez, R. 2006b. Análisis del Viento en el Valle del Río Negro Mediante Mapas Auto Organizados y Algoritmos de Inducción. Avances en Energías Renovables y Ambiente Vol. 10. Pág. 11.137-11.144.
- Cooley, R. 2003. The Use of Web Structure and Content to Identify Subjectively Interesting Web Usage Patterns. ACM Transactions on Internet Technology, 32: 93-116.
- Curtis, B., Kellner, M., Over, J. 1992. Process Modelling. Communications of the ACM, 359: 75-90.
- D'Atri, M., Rodríguez, D., García-Martínez, R. 2009. Improving Pipeline Risk Models by Using Data Mining Techniques. 24th World Gas Conference Proceedings CD. Paper 663.
- Diez, E., Pytel, P., Rodríguez, D., García, R., Lacabanne, M., Leonardis, L., Martins, S., Cartanilica, A., García-Martínez, R. 2012. Aseguramiento de la Calidad para Proyectos de Explotación de Información. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 212-216. ISBN 978-950-766-082-5.
- Fayad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurudsam, R. 1996. Advances in Knowledge Discovery and Data Mining. AAAI Press.
- Felgaer, P., Britos, P. and García-Martínez, R. 2006. Prediction in Health Domain Using Bayesian Network Optimization Based on Induction Learning Techniques. International Journal of Modern Physics C 173: 447-455.
- Ferreira, J., Takai, O., Pu, C. 2005. Integration of Business Processes with Autonomous Information Systems: A Case Study in Government Services. Proceedings Seventh IEEE International Conference on E-Commerce Technology. Pág. 471-474.
- Ferrero, G., Britos, P., García-Martínez, R., 2006. Detection of Breast Lesions in Medical Digital Imaging Using Neural Networks. In IFIP International Federation for Information Processing, Volume 218, Professional Practice in Artificial Intelligence, eds. J. Debenham, Boston: Springer, Pág. 1-10.

- Flores, D., García-Martínez, R., Fernandez, E., Merlino, H., Rodríguez, D., Britos, P. 2009. Detección de Patrones para la Prevención de Daños y/o Averías en la Industria Automotriz. *Proceedings XV Congreso Argentino de Ciencias de la Computación. Workshop de Base de Datos y Minería de Datos*. Págs. 1021-1030. ISBN 978-897-24068-4-1.
- García Martínez, R. 1997. *Sistemas Autónomos. Aprendizaje Automático*. Editorial Nueva Librería. ISBN 950-9088-84-6. García Martínez, R. y Britos, P. 2004. *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería. ISBN 987-1104-15-4.
- García Martínez, R., Servente, M. y Pasquini, D. 2003. *Sistemas Inteligentes*. Editorial Nueva Librería. Buenos Aires. ISBN 987-1104-05-7.
- García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell, J. 2011a. Towards an Information Mining Engineering. En *Software Engineering, Methods, Modeling and Teaching*. Sello Editorial Universidad de Medellín. ISBN 978-958-8692-32-6. Páginas 83-99.
- García-Martínez, R., Britos, P., Pollo-Cattaneo, F., Rodríguez, D., Pytel, P. 2011b. Information Mining Processes Based on Intelligent Systems. *Proceedings of II International Congress on Computer Science and Informatics INFONOR-CHILE 2011*. Pp. 87-94. ISBN 978-956-7701-03-2.
- García-Martínez, R., Lelli, R., Merlino, H., Cornachia, L., Rodríguez, D., Pytel, P., Arboleya, H. 2011c. Ingeniería de Proyectos de Explotación de Información para PYMES. *Proceedings XIII Workshop de Investigadores en Ciencias de la Computación*. Artículo 3761.
- Grossam, 1999. Grossman, R., Kasif, S., Moore, R., Rocke, D., Ullman, J. 1998. Data Mining Research: Opportunities and Challenges. <http://www.csl.mtu.edu/cs5811/common/2002-fall-reading-group/2002-10-25/grossman98.ps>. Ultimo acceso 17 de Abril del 2008.
- Grosser, H., Britos, P. y García-Martínez, R. 2005. Detecting Fraud in Mobile Telephony Using Neural Networks. *Lecture Notes in Artificial Intelligence* 3533: 613-615.
- Jiménez Rey, E., Rodríguez, D., Britos, P., García-Martínez, R. 2009. Caracterización de Problemas de Aprendizaje Basada en Explotación de Información. *Proceedings XI Workshop de Investigadores en Ciencias de la Computación*. Pág. 627-629. ISBN 978-950-605-570-7.
- Kanungo, S. 2005. Using Process Theory to Analyze Direct and Indirect Value-Drivers of Information Systems. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. Pág. 231-240.
- Kuna, H., Pautsch, G., Rey, M., Cuba, C., Rambo, A., Caballero, S., García-Martínez, R., Villatoro, F. 2012. Comparación de la Efectividad de Procedimientos de la Explotación de Información para la Identificación de Outliers en Bases de Datos. *Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación*. Pág. 296-300. ISBN 978-950-766-082-5.
- Kuna, H., Caballero, S., Rambo, A., Meinel, E., Steinhilber, A., Pautsch, G., García-Martínez, R., Villatoro, F. 2010a. Avances en Procedimientos de la Explotación de Información Para la Identificación de Datos Faltantes, con Ruido e Inconsistentes. *Proceedings XII Workshop de Investigadores en Ciencias de la Computación*. Pág. 137-141.
- Kuna, H., Caballero, S., Rambo, A., Meinel, E., Steinhilber, A., Pautsch, G., Rodríguez, D., García-Martínez, R., Villatoro, F. 2010b. Identification of Noisy Data in Databases by Means of a Clustering Process. En *Ingeniería de Software e Ingeniería del Conocimiento: Tendencias de Investigación e Innovación Tecnológica en Iberoamérica* Editores: R. Aguilar, J. Díaz, G. Gómez, E- León. Pág. 264-273. Alfaomega Grupo Editor. ISBN 978-607-707-096-2.
- Kuna, H., García Martínez, R., Villatoro, F. 2010c. Pattern Discovery in University Students Desertion Based on Data Mining. *Advances and Applications in Statistical Sciences Journal*, 22: 275-286. ISSN 0974-6811.
- Kuna, H., García Martínez, R., Villatoro, F. 2009. Identificación de Causales de Abandono de Estudios Universitarios. *Uso de Procesos de Explotación de Información*. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología 5: 39-44.
- Kuna, H., Pautsch, G., Rey, M., Cuba, C., Rambo, A., Caballero, S., Steinhilber, A., García-Martínez, R., Villatoro, F. 2011. Avances en Procedimientos de la Explotación de Información con Algoritmos basados en la Densidad para la Identificación de Outliers en Bases de Datos. *Proceedings XIII Workshop de Investigadores en Ciencias de la Computación*. Artículo 3745.
- Langseth, J., Vivatrat, N. 2003. Why Proactive Business Intelligence is a Hallmark of the Real-Time Enterprise: Outward Bound. *Intelligent Enterprise* 518: 34-41.
- Lopez-Nocera, M., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2011. Un Protocolo de Caracterización Empírica de Dominios para Uso en Explotación de Información. *Proceedings XVII Congreso Argentino de Ciencias de la Computación*. Pág. 1047-1055. ISBN 978-950-34-0756-1.
- Lyman, P., Varain, H. 2003. How Much Information?. *School of Information Management & Systems*. University of California Berkeley. <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>. Ultimo acceso 21 de Mayo del 2008.
- Maimon, O., Rokach, L. 2005. *The Data Mining and Knowledge Discovery Handbook*. Springer Science + Business Media Publishers.
- Mansilla, D., Pollo-Cattaneo, M., Pytel, P., García-Martínez, R. 2012. Modelo de Proceso para Elicitación de Requerimientos en Proyectos de Explotación de Información. *Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación*. Pág. 222-225. ISBN 978-950-766-082-5.
- Marbán, O. 2003. *Modelo Matemático Paramétrico de Estimación para Proyectos de Data Mining (DMCoMo)*. Tesis de Doctorado en Informática. Universidad Politécnica de Madrid.
- Mobasher, B., R Cooley and J Srivastava 1999. Creating adaptive web sites through usage-based clustering of URLs. *Proceedings Workshop on Knowledge and Data Engineering Exchange*, Pág. 19-25.
- Moss, L. 2003. Nontechnical Infrastructure of BI Applications. *DM Review* 131: 42-45.
- Negash, S., Gray, P. 2008. Business Intelligence. En *Handbook on Decision Support Systems* 2, ed. F. Burstein y C. Holsapple Heidelberg, Springer, Pág. 175-193.
- Ochoa, A., Fernández, E., Britos, P., García-Martínez, R. 2008. *Metodologías de Ingeniería Informática*. Editorial Nueva Librería. ISBN 978-987-1104-54-3.
- Pollo-Cattaneo, F., Britos, P., Pesado, P., García-Martínez, R. 2010a. Ingeniería de Procesos de Explotación de Información. En *Ingeniería de Software e Ingeniería del Conocimiento: Tendencias de Investigación e Innovación Tecnológica en Iberoamérica* Editores: R. Aguilar, J. Díaz, G. Gómez, E- León. Pág. 252-263. Alfaomega Grupo Editor. ISBN 978-607-707-096-2.
- Pollo-Cattaneo, F., Britos, P., Pesado, P., García-Martínez, R. 2010b. Proceso de Educación de Requisitos en Proyectos de Explotación de Información. En *Ingeniería de Software e Ingeniería del Conocimiento: Tendencias de Investigación e Innovación Tecnológica en Iberoamérica* Editores: R. Aguilar, J. Díaz, G. Gómez, E- León. Pág. 01-11. Alfaomega Grupo Editor. ISBN 978-607-707-096-2.
- Pollo-Cattaneo, M., García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Rodríguez, D., Merlino, H., Pytel, P., Vanrell, J. 2012a. Elementos para una Ingeniería de Explotación de Información. *Proyecciones* 101: 67-84. ISSN 1667-8400.
- Pollo-Cattaneo, M., García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Rodríguez, D., Merlino, H., Pytel, P., Vanrell, J. 2012b. Elementos para una Ingeniería de Explotación de Información. *Proyecciones* 101: 67-84. ISSN 1667-8400.
- Pyle, D. 2003. *Business Modeling and Business intelligence*. Morgan Kaufmann Publishers.
- Pytel, P., Britos, P., García-Martínez, R. 2012c. Viabilidad y Estimación de Proyectos de Explotación de Información. *Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación*. Pág. 217-221. ISBN 978-950-766-082-5.
- Pytel, P., Pollo-Cattaneo, F., Rodríguez, D., Britos, P., García-Martínez, R. 2011b. Identificación de Tareas Críticas en una Metodología de Desarrollo de Proyectos de Explotación. *Proceedings XVII Congreso Argentino de Ciencias de la Computación*. Pág. 989-998. ISBN 978-950-34-0756-1.
- Pytel, P., Tomasello, M., Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2011a. Estudio del Modelo Paramétrico DMCoMo de Estimación de Proyectos de Explotación de Información. *Proceedings XVII Congreso Argentino de Ciencias de la Computación*. Pág. 979-988. ISBN 978-950-34-0756-1.
- Pytel, P., Tomasello, M., Rodríguez, D., Arboleya, H., Pollo-Cattaneo, M., Britos, P., García-Martínez, R. 2011c. Estimación de Proyectos de Explotación de Información. *Estudio Comparado de Modelos Analíticos y Empíricos*. *Proceedings XIII Workshop de Investigadores en Ciencias de la Computación*. Artículo 3848.
- Rancan, C., Pesado, P., García-Martínez, R. 2010. Issues in Rule Based Knowledge Discovering Process. *Advances and Applications in Statistical Sciences Journal* ISSN 0974-6811, 22: 303-314. ISSN 0974-6811.
- Rodríguez, D., Pollo-Cattaneo, F., Britos, P., García-Martínez, R. 2010. Estimación Empírica de Carga de Trabajo en Proyectos de Explotación de Información. *Anales del XVI Congreso Argentino de Ciencias de la Computación*. Pág. 664-673. ISBN 978-950-9474-49-9.
- Rudin, K., Cressy, D. 2003. Will the Real Analytic Application Please Stand Up? *DM Review* 133: 30-34.
- Saavedra-Martínez, P., Pollo-Cattaneo, F., Rodríguez, D., Britos, P., García-Martínez, R. 2012a. Proceso de Identificación de Errores de Apropriación de Conceptos Basado en Explotación de Información. *Proceedings VII Congreso de Tecnología en Educación y Educación en Tecnología*. ISBN 978-987-28186-0-9. Universidad Nacional del Noroeste de Buenos Aires. Pergamino. Buenos Aires. Argentina.
- Saavedra-Martínez, P., Pollo-Cattaneo, M., Pytel, P., Rodríguez, D., García-Martínez, R. 2012b. Detección de Problemas de Aprendizaje Basado en Explotación de Información. *Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación*. Pág. 962-964. ISBN 978-950-766-082-5.
- Sanson, E., Britos, P., Rodríguez, D., García-Martínez, R. 2009. Clasificación Automática para la Prevención del Estrés de los Suelos y la Fatiga de Soja en el Noroeste Argentino. *Proceedings XI Workshop de Investigadores en Ciencias de la Computación*. Pág. 333-335. ISBN 978-950-605-570-7.
- SAS, 2008. SAS Enterprise Miner: SEMMA. <http://www.sas.com/technologies/analytics/datamining/miner/semma.html>. Ultimo acceso Junio 2008.
- Srivastava, J., Cooley, R., Deshpande, M., Tan, P. 2000. Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. *SIGKDD Explorations*, 12: 12-23.
- Thomsen, E. 2003. BI's Promised Land. *Intelligent Enterprise*, 64: 21-25.
- Valenga, F., Fernández, E., Merlino, H., Rodríguez, D., Procopio, C., Britos, P., García-Martínez, R. 2008. Minería de Datos Aplicada a la Detección de Patrones Delictivos en Argentina. *Proceedings VII Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento*. Pág. 31-39. ISSN 1390-292X.
- Valenga, F., Perversi, I., Fernández, E., Merlino, H., Rodríguez, D., Britos, P., García-Martínez, R. 2007a. Estudio Preliminar: La Estadística Criminal y el Aporte de la Minería de Datos. En Kaminsky, G., Kosovsky, D., Kessler, G. *El Delito en la Argentina Post-crisis*. pp 11-24. Editado por la Friedrich Ebert Stiftung.
- Valenga, F., Perversi, I., Fernández, E., Merlino, H., Rodríguez, D., Britos, P., García-Martínez, R. 2007b. Aplicación de Minería de Datos para la Exploración y Detección de Patrones Delictivos en Argentina. *Anales del XIII Congreso Argentino de Ciencias de la Computación*. Pág. 258-270. ISBN 978-950-656-109-3.
- Vanrell, J., Bertone, R., García-Martínez, R. 2010a. Modelo de Proceso de Operación para Proyectos de Explotación de Información. *Anales del XVI Congreso Argentino de Ciencias de la Computación*. Pág. 674-682. ISBN 978-950-9474-49-9.
- Vanrell, J., Bertone, R., García-Martínez, R. 2010b. Un Modelo de Procesos de Explotación de Información. *Proceedings XII Workshop de Investigadores en Ciencias de la Computación*. Pág. 167-171.
- Vegega, C., Mansilla, D., Pollo-Cattaneo, M. F., Pytel, P., Rodríguez, D., Diez, E., García-Martínez, R. 2012. Documentación de Requisitos en Proyectos de Explotación de Información. *Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación*. Pág. 226-229. ISBN 978-950-766-082-5.
- Vuori, V. 2006. The Employees as a Source of External Business Information. *Proceedings European Productivity Conference EPC'06*. Pág. 29-36.